

Physiological Measurement



PAPER • OPEN ACCESS

Motor execution detection based on autonomic nervous system responses

To cite this article: Laura Marchal-Crespo *et al* 2013 *Physiol. Meas.* **34** 35

View the [article online](#) for updates and enhancements.

Related content

- [Peripheral autonomic signals for body-machine interface control](#)
Tiago H Falk, Mirna Guirgis, Sarah Power *et al*.
- [Classification of prefrontal activity due to mental arithmetic and music imagery](#)
Sarah D Power, Tiago H Falk and Tom Chau
- [Rehabilitation of hand in subacute tetraplegic patients based on brain computer interface and functional electrical stimulation: a randomised pilot study](#)
Bethel C A Osuagwu, Leslie Wallace, Mathew Fraser *et al*.

Recent citations

- [Luz Maria Alonso-Valerdi and Andrés Antonio González-Garrido](#)
- [Classification of Anticipatory Signals for Grasp and Release from Surface Electromyography](#)
Ho Siu *et al*
- [Frequency-response-based analysis of respiratory sensor measuring capacitance built across skin](#)
Makie Terasawa *et al*

Motor execution detection based on autonomic nervous system responses

Laura Marchal-Crespo^{1,2}, Raphael Zimmermann^{3,4},
Olivier Lambercy³, Janis Edelmann³, Marie-Christine Fluet³,
Martin Wolf⁴, Roger Gassert³ and Robert Riener^{1,2}

¹ Sensory-Motor Systems (SMS) Laboratory, Institute of Robotics and Intelligent Systems (IRIS), ETH Zurich, 8006 Zurich, Switzerland

² Medical Faculty, Balgrist University Hospital, University of Zurich, Switzerland

³ Rehabilitation Engineering Laboratory, ETH Zurich, 8006 Zurich, Switzerland

⁴ Biomedical Optics Research Laboratory, University Hospital Zurich, University of Zurich, 8006 Zurich, Switzerland

E-mail: laura.marchal@hest.ethz.ch

Received 17 September 2012, accepted for publication 15 November 2012

Published 18 December 2012

Online at stacks.iop.org/PM/34/35

Abstract

Triggered assistance has been shown to be a successful robotic strategy for provoking motor plasticity, probably because it requires neurologic patients' active participation to initiate a movement involving their impaired limb. Triggered assistance, however, requires sufficient residual motor control to activate the trigger and, thus, is not applicable to individuals with severe neurologic injuries. In these situations, brain and body–computer interfaces have emerged as promising solutions to control robotic devices. In this paper, we investigate the feasibility of a body–machine interface to detect motion execution only monitoring the autonomic nervous system (ANS) response. Four physiological signals were measured (blood pressure, breathing rate, skin conductance response and heart rate) during an isometric pinching task and used to train a classifier based on hidden Markov models. We performed an experiment with six healthy subjects to test the effectiveness of the classifier to detect rest and active pinching periods. The results showed that the movement execution can be accurately classified based only on peripheral autonomic signals, with an accuracy level of 84.5%, sensitivity of 83.8% and specificity of 85.2%. These results are encouraging to perform further research on the use of the ANS response in body–machine interfaces.

Keywords: body–machine interface, autonomic nervous system, physiological measurements, hidden Markov model, robot-assisted rehabilitation

(Some figures may appear in colour only in the online journal)



Content from this work may be used under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 licence](https://creativecommons.org/licenses/by-nc-sa/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

1. Introduction

There is increasing interest in using robotic devices to assist individuals who suffered neurologic injuries such as stroke and spinal cord injury (Marchal-Crespo and Reinkensmeyer 2009). Neurologic patients' active participation is thought to be essential for provoking motor plasticity (Lotze *et al* 2003, Perez *et al* 2004), and by assisting the movement that participants cannot achieve by themselves, active assist exercise provides novel somatosensory stimulation that can help induce brain plasticity (Rossini and Dal Forno 2004). Triggered assistance allows the participant to attempt a movement without any robotic assistance, only initiating the assistance when some performance variable (e.g. force generated by the participant, limb velocity, or muscle activity measured with surface EMG) reaches a threshold.

Triggered robotic assistance, however, requires sufficient residual motor ability or remaining muscle activity to activate the trigger, and hence is not applicable to individuals who have no functional motor ability left as a result of a severe neurologic injury. In these situations, brain–computer interfaces (BCI) have emerged as promising solutions (del R Millan *et al* 2010). Brain–computer interfaces could be used to control robotic devices to move the impaired limb when an intention to move is detected from cortical activity. Intention to move is defined as a supraspinal command that results in a physiological change, and eventually in a movement. Electroencephalography (EEG) and, more recently, functional near-infrared spectroscopy (fNIRS) are the most widely used non-invasive techniques employed in BCIs. However, the burden of connecting sensors on the patients scalp and the relatively long training period required for the user to produce classifiable brain signals can be time consuming and frustrating. Additionally, the system performance can be severely affected by the interference caused by sensor location and, in the case of fNIRS, hair color and thickness. All these challenges can lead to user frustration and, ultimately, rehabilitation withdrawal (Coyle *et al* 2004, van Gerven *et al* 2009).

More recently, studies have introduced the concept of body–machine interfaces (BMI), where physiological signals can be self controlled and used to detect functional intent (see Blain *et al* (2008) for a review). Responses of the autonomic nervous system (ANS), such as cardiorespiratory and electrodermal responses, can be measured with economical off-the-shelf instrumentation and are relatively fast to set up. Physiological signals such as skin conductance response, heart rate, respiration rate and skin temperature have been shown to have the potential of serving as inputs for the development of BMIs (Blain *et al* 2008). However, these previous studies in BMIs are mainly based on self-paced physiological signal changes, and thus the approach still requires the subject to perform a training phase to learn how to successfully control his/her physiological signals. Nevertheless, recent studies in psychophysiology showed that non-self-paced physiological signals can also provide a proper method to estimate a person's emotion and frustration level (Kim and Andre 2008, Scheirer *et al* 2002), mental workload (Wilson and Russell 2003, Collet *et al* 2009) and activity engagement (Kushki *et al* 2012) without his/her active participation.

Physiological measurements have also been employed to increase the performance of BCIs. These so-called hybrid BCIs use brain recording technologies in conjunction to physiological signals (e.g. heart rate and blood pressure) to improve the classification performance (for a review, see Pfurtscheller *et al* (2010)). Most of the work on hybrid BCIs has made use of self-paced physiological signals, whereas there are only few studies that employed non-self-paced physiological data. An example of non-self-paced hybrid BCIs that outperformed classic BCIs included the respiration rate, heart rate, skin temperature and skin conductance response in an BCI based on music imagery (Falk *et al* 2011). We recently conducted an experiment, the results of which showed that the addition of blood

pressure, respiration rate, heart rate, and skin conductance response significantly improved the accuracy of detecting motor execution of an fNIRS-based BCI (Zimmermann *et al* 2012). Interestingly, while hybrid BCIs have been proposed as an alternative to classic BCIs to improve accuracy, physiological signals have never been employed as stand-alone signals to detect motion execution. This paper suggests a paradigm shift into the use of the ANS responses in BCIs: the physiological signals are treated as the unique main source of information.

This paper investigates the feasibility of a BMI to detect motor execution, monitoring only changes in peripheral autonomic signals, without direct measurement of force, EMG activity and brain activation. The motivation behind our approach is to provide an interface for severely affected neurological patients who cannot rely on their neural circuitry to trigger assistance or control a robotic device. We hypothesize that such a BMI can achieve similar performance in detecting motor execution as BCIs directly based on signals from the central nervous system. This technology could improve not only robot-assisted rehabilitation, but also assist during activities of daily living: a mobile robot or a wearable exoskeleton in a home environment could provide support during any task based on subject's motion intention.

We performed an experiment with six healthy subjects. Four physiological signals were acquired (mean blood pressure, breathing rate, skin conductance response and heart rate) during an isometric pinching task. The physiological signals were used to train and evaluate an individually optimized classifier to detect rest and active pinching periods based on hidden Markov models (HMMs). The rationale behind an individually optimized classifier, rather than the one that generalizes to a wide range of users, is to study the feasibility of a classifier that could ultimately tune its parameters to different subjects, e.g., subjects with neurological injuries such as stroke or SCI.

2. Methods

2.1. Measurements of physiological responses

Based on previous research in the fields of BMI and psychophysiology (Blain *et al* 2008, Koenig *et al* 2011), four peripheral autonomic signals were recorded online: electrocardiogram (ECG), respiration, blood pressure, and skin conductance response (SCR). All physiological signals were acquired at 600 Hz using a biosignal amplifier (g.USBamp, g.tec, Austria, figure 1).

2.1.1. ECG. ECG was measured using the g[®].GAMMAsys active electrode system from g.tech. The electrodes (g[®].GAMMAclip, g.tec, Austria) were placed using sticky patches, with the ground on the left shoulder, reference over the left clavicle, channel 1 over the right ribs and channel 2 over the left ribs. The skin area where electrodes were placed was previously cleaned, although no further skin preparation was required (i.e. shaving).

The raw ECG signal was filtered with a fourth-order Butterworth bandpass filter with the frequency band 0.01–40 Hz. The heart rate (HR) was calculated online, detecting the R-wave peaks of the QRS complex using an adaptive threshold algorithm similar to the one described by Christov (2004). The HR was simultaneously calculated using a similar adaptive threshold on the raw blood pressure signal and compared to the HR calculated from the ECG in order to increase the HR detection robustness. Time and frequency domain measures of heart rate variability were discarded as possible features, since the minimum time interval required to measure cardiovascular variability is typically 5 min (Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology 1996).



Figure 1. Measurement setup. Four physiological signals were acquired: blood pressure, respiration rate, skin conductance response and electroencephalogram.

2.1.2. Respiration rate. The respiration signal was acquired using a thermistor respiration flow sensor (SleepSense®, Scientific Laboratory Products, USA) placed at the entrance of the nostrils. The sensor was fixed on the skin using hypoallergenic adhesive tape.

The raw respiration signal, measured as the difference of temperature between inhaled and exhaled air, was filtered with an eighth-order Butterworth bandpass filter with the frequency band 0.1–2.1 Hz. The breathing rate (BR) was calculated using an adaptive threshold algorithm, similar to the one employed with ECG. Time and frequency domain measures of breathing variability were not considered due to the short recording periods. Other respiration-related measurements, such as breathing amplitude, were excluded after preliminary testing, since no changes between rest and active periods were observed.

2.1.3. Blood pressure. The raw blood pressure was measured with a continuous non-invasive arterial pressure system (CNAP™ monitor 500, CNSystems, Austria). An inflatable cuff was placed around the left upper arm, and two size-adjustable finger cuffs were attached to the proximal phalanges of the left index and middle fingers. Subjects were requested to position the left arm on the chest over the heart. The arm cuff was employed only for a couple of minutes during the system initialization for scaling purposes. During the experiments, only the finger cuffs were used.

The raw blood pressure signal was detrended subtracting a best-fit line (in the least-squares sense) from the raw signal in order to remove any possible signal drifts during sessions. It was further low-pass filtered with a first-order Butterworth filter with the cutoff frequency of 0.1 Hz, leaving only the low and very low frequency spectra of the signal. The selection of the low and very low spectra (called mean blood pressure in subsequent sections, BP) was performed after comparing which cardiovascular features showed the most significant changes between rest and activation periods. Thus, diastolic, systolic and raw blood pressure signals, although initially considered, were excluded after preliminary testing.

2.1.4. Skin conductance response. Skin conductance was measured attaching two electrodes (g®.GSRsensor, g.tec, Austria) through Velcro® rings to the distal phalanges of the left index

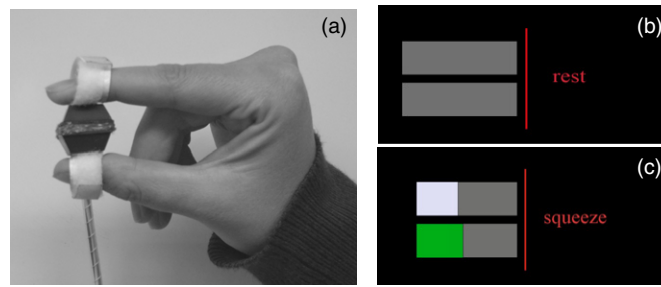


Figure 2. (a) Force sensor attached to the distal phalanges of the index finger and thumb of the right hand with Velcro® rings. (b) Visual command Rest. (c) Visual command activity. The reference force is rendered as a dynamic green bar (lower bar), while the real-time measured force was visually represented by a dynamic white bar.

and middle fingers. Skin conductance is characterized by a slowly changing background level (tonic), and a rapid time-varying (phasic) response (Malmivuo and Plonsey 1995). The tonic level is related to the general activity of the perspiratory glands influenced by external temperature. The phasic response is called SCR and is usually related to the automatic response to stimuli.

The raw skin conductance signal was filtered with an eighth-order Butterworth low-pass filter with a cutoff frequency of 30 Hz. The SCR signal was linearly detrended over each rest-activity period to remove the tonic level and was further normalized.

2.2. Experimental protocol

All experiments were approved by the institutional ethics committee of the ETH Zurich (application number EK 2010-N-49), and participants were provided informed consent. Six healthy male subjects between the age of 20–30 yr were recruited from the ETH Zurich students and staff environment. Inclusion criteria were no history of neurological disorders or orthopedic problem affecting the right upper extremity.

The measurements were conducted in a silent, dark room. Subjects were requested to lie supine on a comfortable padded table. The task consisted in isometrically tracking a provided pinching force with the right index finger and thumb. Isometric pinching was chosen partly for convenience (i.e. it is a simple task that minimizes subject movement), but also because it allows for a systematic assessment of subject's performance. The force applied by the subject during pinching was measured with a one-axis thick-film force sensor (CentoNewton 100N, LPM-EPFL, Switzerland) attached to the distal phalanges with Velcro® rings (figure 2(a)). Subjects were instructed to remain as motionless as possible during the experiment. The protocol was implemented in Simulink®. The force sensor was connected to the computer via a USB data acquisition card (NI USB-6008, National Instruments Inc., USA).

The experimental protocol was described in detail by Zimmermann *et al* (2011). Here, only a brief summary is given for completeness. fNIRS was used to simultaneously record brain activity in motor areas; however, these signals are not used in the present analysis and are beyond the scope of this paper. Three visual commands were presented to subjects using video goggles (z920HR-VGA, Zetronix Corp., USA): (1) rest: the word *rest* was displayed on the screen (figure 2(b)), and subjects were instructed to remain as relaxed as possible; (2) preparation: the message *get ready* was displayed and subjects were instructed to be

aware they will be asked to move in a few seconds. Subjects were not instructed to imagine the movement or to try to move as quickly as possible; (3) activity: the word *squeeze* was displayed (figure 2(c)). Subjects were requested to try to match their applied pinching force (visually represented by a dynamic horizontal white bar, figure 2(c)) with a reference force (rendered with a horizontal green bar under the reference bar). In order to prevent subjects from learning the reference force and reduce their concentration level, a complex reference force profile between 1 and 4 N was generated from a truncated Fourier series with frequencies 0.5, 1.0 and 1.1 Hz. The force level and duration of the activity periods were small enough to avoid fatigue.

The protocol consisted of a random presentation of three different sequences of visual commands.

- S1 and S2: a rest command was followed by a preparation command (of 10 or 5 s), and then followed by an activity command that lasted for 20 s.
- S3: a rest command was followed by a preparation command of 10 s, and followed again by a rest command.
- S4: a rest command was followed by an activity command that lasted for 20 s.

The duration of the rest commands was randomized (from 15 to 24 s) in order to reduce learning effects that may decrease attention and to avoid that the autonomic system synchronizes with the activity periods. The experimental protocol consisted of a total of 10 trials per sequence, presented in random order. The experiment was divided into two sessions of 20 trials each. Each session began with a baseline of 180 s and finished with a baseline of 120 s. The total time required to finish a session was approximately 20 min. Participants paused for 10 min between sessions.

2.3. Classifier

In a preliminary study (Zimmermann *et al* 2011), we found that none of the physiological signals showed a significant change between the rest and the preparation periods. Thus, only periods of activity and the rest periods that preceded were considered in the classifier, independently of the sequences they were part of (i.e. 10 trials per sequences S1, S2 and S4, and thus, a total of 30 rest periods followed by 30 activity periods).

2.3.1. Data pre-processing. The four physiological signals were further decimated to 5 Hz in order to reduce the computational time required to train and evaluate the classifier. The training and testing data sets were generated as vectors of physiological values at each sample time. A 15 s window was selected, corresponding to the shortest rest command possible. Optimizing the window length for each subject would increase considerably the training time of the classifier, and thus, the same conservative window length was fixed for all subjects. Responses of the ANS are rather slow (figure 3), and thus choosing the last 15 s of the rest periods, and the first 15 s of the active periods did not seem reasonable. Different physiological signals have different latency responses (i.e. the SCR generally shows a faster response than other systemic changes, figure 3), and thus, different post-stimulus times could ideally be used for each signal to detect the active periods. However, in order to reduce the computational time that optimizing the latency time for each subject and for each physiological signal would require, we fixed the latency time to 5 s, and thus, the 15 s windows were shifted ahead by 5 s.

A fourfold cross-validation was used to randomly distribute all pairs of associated rest-pinching trials into train and test data sets (i.e. 25 training trials and 5 testing trials).

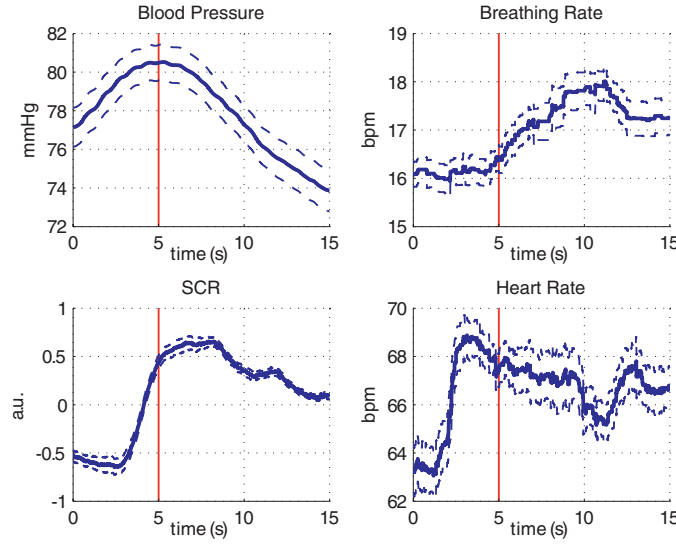


Figure 3. Mean values (solid) and standard error (dashed) of the biosignals during the first 15 s of the 30 pinching periods for subject 3. The red vertical line represents the ad hoc decided latency.

2.3.2. Hidden Markov models. HMMs are well known in temporal pattern recognition applications such as speech and gesture recognition. The main argument for HMMs over other classification techniques (e.g. support vector machines, linear discriminant analysis) is their ability to classify time-sequential data, such as the time-varying physiological signals presented. Here, only a brief introduction to HMMs is given. The reader is referred to Rabiner (1989) for a detailed tutorial.

A HMM is a finite-state machine containing N unobservable (hidden) states ($S = \{S_1, S_2, \dots, S_N\}$). The probability of transition to other states only depends on the current state and is defined by a transition probability matrix $A = [a_{11}a_{12}\dots; a_{21}\dots; \dots a_{NN}]$. HMMs emit an observation vector at every time sample $O_t = \{O_1, O_2, \dots, O_F\}$ that depends only on the current state and number of features F . Each state has an associated observation probability distribution B which determines the probability of generating an observation at a certain time step. The probability of starting in a specific state is modeled by the initial state distribution π .

A HMM is completely characterized by defining the number of states N , the initial and transition state probabilities (π and A), and the observation probability distributions B at each state (denoted in short as $\lambda(\pi, A, B)$). In this paper, a left-right Markov model topology was chosen that allowed transitions only from each state to itself and to the state to its right (figure 4). The observation probability distributions were chosen to be mixtures of M Gaussians with full covariance matrices in order to account for possible observation correlations. To reduce the chance of overfitting the classifier, only HMMs with a maximum of five states and two mixtures were considered. The number of observations was set to 4 described in section 2.1 (figure 3).

The initial transition matrix A_0 and initial state probability π_0 were estimated by uniformly distributed random numbers. The observation probability distribution B_0 was initialized using k -means clustering on training observations. In order to find the optimal model parameters $\lambda(\pi, A, B)$ given a fixed number of states and mixtures, the initial probability parameters were

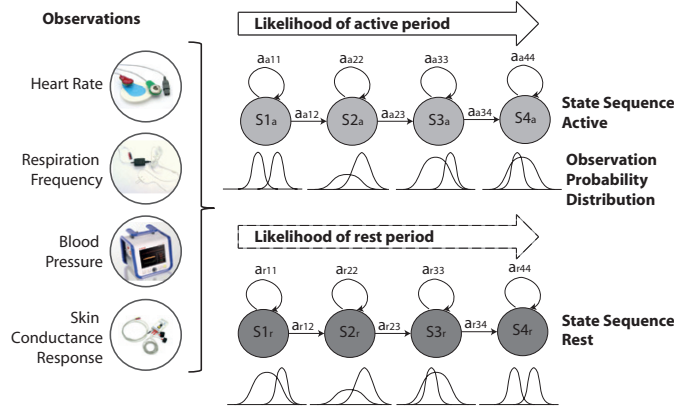


Figure 4. Illustration of two four-state left-right HMMs for rest and active conditions. The four observations employed in this study were: heart rate (HR), respiratory frequency (RF), mean blood pressure (BP) and skin conductance response (SCR). The observation probability distributions at each state are represented as mixtures of two Gaussians.

adjusted using the Baum–Welch algorithm (Rabiner 1989) on the training data. The freely distributed HMM toolbox for Matlab by Murphy (1998) was used.

Given a specific number of states and mixtures and a sequence of test observations, the likelihood that the observed sequence was produced by any of the two HMM models (one for rest, and a second one for active) was computed, using the forward–backward algorithm (Rabiner 1989). Subsequently, each of the testing trials was classified into one of the two models, by selecting the model with the highest likelihood. Based on the different number of hidden states $N = \{1, 2, 3, 4, 5\}$ and different number of mixtures $M = \{1, 2\}$, a total of ten models were trained for each of the rest and active classes.

2.4. Evaluation

The metrics used to quantify the classifier performance were accuracy, sensitivity and specificity:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100, \quad (1)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100, \quad (2)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100, \quad (3)$$

where TP is the number of true positives (correctly detected active periods), TN is the number of true negatives (correctly detected rest periods), FP is the number of false positives (rest periods classified erroneously as active) and FN is the number of false negative (active periods classified as rest). The performance metrics were calculated for each k -fold partition and averaged over one complete cross-validation run.

A well-known problem with HMMs is their lack of convergence to a global maximum. Changing the initial model parameters $\lambda(\pi_0, A_0, B_0)$ results in a different optimized trained model. In order to reduce the effect on performance variability due to random initial model parameters, we ran the evaluation procedure a total of 7 times. The mean and the standard

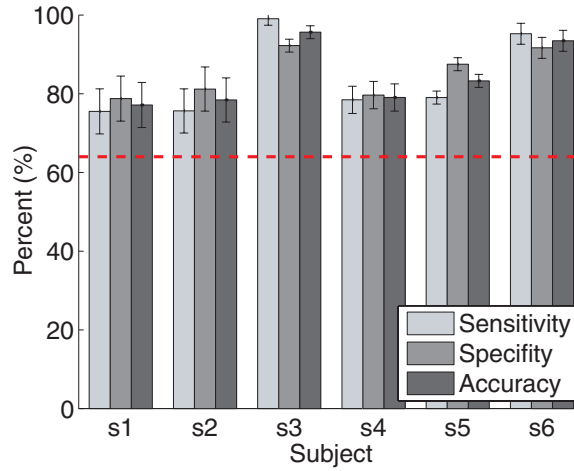


Figure 5. Mean and SD classification sensitivity, specificity and accuracy across seven complete cross-validation runs for each subject, and for the best combination of number of states N and number of mixtures M . The dashed line indicates the level of 64.1% considered as the value above which the performance variables are above the chance level (66.5% for subject 6).

deviation (SD) of the performance metrics based on the seven completed cross-validation runs were computed.

Each subject showed different biosignal response profiles, and thus, the HMM configurations (i.e. number of states and mixtures) were optimized for each subject. The model topology with the highest mean accuracy was selected from the total of 10 possible model structures (number of hidden states $N = \{1, 2, 3, 4, 5\}$ and number of mixtures $M = \{1, 2\}$).

The chance level in a two-class BCI is not exactly 50%, but 50% with a confidence interval at a certain level (95%) that depends on the number of training trials (Muller-Putz *et al* 2008). The calculation of the confidence interval was performed using a binomial distribution of $p = 0.5$, considering 30 trials per class (two sessions), and 15 trials per class (one session). This yielded upper confidence limits of 64.1% when considering the two sessions and 66.5% for one session. Thus, the obtained performance metrics were considered above the chance level when their means were significantly higher than the corresponding upper confidence limits. The significant level was set to $p = 0.05$.

3. Results

Due to technical problems, the physiological signals of subject 6 were recorded only during the first session. Figure 5 reports the per-participant classifier performance sensitivity, specificity and accuracy values obtained using the four features with the combination of number of states N and number of mixtures M that yielded the maximum accuracy. The optimum number of states and mixtures per subject, and the mean and SD of the sensitivity, specificity and accuracy for each subject are reported in table 1. All subjects performed significantly above the a priori set 64.1% chance threshold (66.5% for subject 6).

The average classifier accuracy over the six participants was 84.5%. The optimization of the HMM parameters (N and M) for each user required a calibration session with known active and rest intervals from the fourfold cross-validation training data set. The calibration phase that iterates for each number of states and mixtures combination takes a relative long time. In practical BCI situations, however, it would be desirable to reduce the time required for

Table 1. Classification sensitivity, specificity and accuracy, (mean \pm SD) across seven complete cross-validation runs for each subject, for the best combination of number of states and number of mixtures, and for a fixed HMM structure ($N = 3$, $M = 2$). The optimum number of states N and mixtures M for each subject in the personalized HMM are also reported.

Subject	Personalized HMM					Fixed HMM		
	Sensitivity (%)	Specificity (%)	Accuracy (%)	N	M	Sensitivity (%)	Specificity (%)	Accuracy (%)
s1	75.5 \pm 7.2	78.8 \pm 9.9	77.1 \pm 5.7	3	2	75.5 \pm 7.2	78.8 \pm 9.9	77.1 \pm 5.7
s2	75.6 \pm 4.0	81.2 \pm 8.5	78.4 \pm 5.6	5	2	76.7 \pm 5.5	79.5 \pm 7.5	78.1 \pm 5.0
s3	99.0 \pm 3.7	92.2 \pm 2.5	95.6 \pm 1.6	2	2	98.0 \pm 2.8	92.9 \pm 3.0	95.4 \pm 2.4
s4	78.4 \pm 1.7	79.7 \pm 7.8	79.1 \pm 3.5	2	1	77.6 \pm 4.7	78.3 \pm 5.7	77.9 \pm 2.8
s5	79.0 \pm 4.3	87.5 \pm 4.7	83.3 \pm 1.7	3	2	79.0 \pm 4.3	87.5 \pm 4.7	83.3 \pm 1.7
s6	95.2 \pm 3.3	91.7 \pm 2.9	93.5 \pm 2.7	4	2	92.6 \pm 2.4	90.5 \pm 5.6	91.5 \pm 3.0
Average	83.8 \pm 3.7	85.2 \pm 6.1	84.5 \pm 3.5			83.2 \pm 4.5	84.6 \pm 6.1	83.9 \pm 3.4

calibration, while still obtaining good accuracies. In order to reduce the calibration computation time during normal BCI investigations, it seems reasonable to use a fixed HMM structure. The effect that fixing the number of states and mixtures had on the overall accuracy was studied. It was found that the number of states and mixtures that maximizes the overall accuracy ($N = 3$ and $M = 2$) resulted in only a slight reduction of 0.6% of the average classifier accuracy (83.9%). The mean and SD of the sensitivity, specificity and accuracy for each subject using a fixed HMM structure are reported in table 1. The accuracies of all subjects remained above the chance level.

Some subjects showed higher accuracy levels than others (i.e. subjects 3, 5 and 6; Mann–Whitney test, $p = 0.05$), probably due to the intersubject differences in the ANS responses. In order to investigate the changes in the different physiological signals, the mean of the last 5 s of the rest periods was compared to the mean value from 5 s during the activity periods. Because different physiological signals have different latency responses (see, e.g., figure 3), different times after the onset of the activity period were used for each signal (3 s post-stimulus for the SCR and 5 s for all the others) (Zimmermann *et al* 2011). Paired t-tests were used to evaluate the presence of a significant change in each physiological signal for each subject. The significance level was set to 5%. The resulting p -values are listed in table 2.

A significant correlation between the accuracy and the number of significant features was found: subjects with a larger number of significant features (e.g. s3 with four significant features, table 2) resulted in higher accuracy (Pearson’s correlation, $R^2 = 0.78$, $p = 0.02$). In order to investigate the effect that non-significant features had on the classifier accuracy, we performed feature reduction (i.e. selection of a subset of features for the classification). There exist several methods for feature reduction. A common technique to rank individual features is through ANOVA: a statistical method used to rank the features which show the most significant difference between two classes. Then, only the n most significant features are used in the classifier (Wagner *et al* 2005). Analysis of variance was chosen to reduce the feature dimensionality because the final extracted features are a subset of the original features, while other popular methods (e.g. principal component analysis) create new transformed features. Furthermore, ANOVA is computationally less expensive when compared to recursive feature reduction algorithms (e.g. sequential forward selection).

Based on the values reported in table 2, the p -values from more to less significant (order shown in brackets) were ranked. For each subject, the feature with the lowest p -value was iteratively added in the features list, and the classifier performance was recalculated. The classification accuracies (mean \pm SD) across the seven complete cross-validation runs for

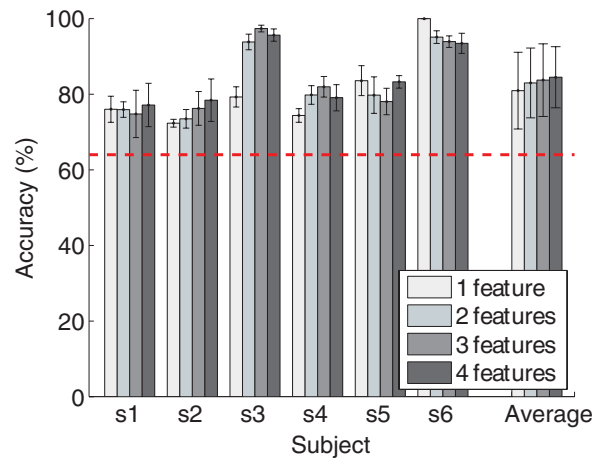


Figure 6. Classification accuracies (%) (mean \pm SD) across seven complete cross-validation runs for each subject and the average of all subjects, for the best combination of number of states and number of mixtures, and for different number of features. The dashed line indicates the level of 64.1% considered as the value above which the performance variables are above the chance level (66.5% for subject 6).

Table 2. *P*-values of variation in physiological signals between rest and active periods for each subject.

Subject	Blood pressure	Breathing rate	SCR	Heart rate
s1	0.1254 (3)	0.0449 ^a (1)	0.0598 (2)	0.9426 (4)
s2	0.0001 ^b (1)	0.6583 (3)	0.0004 ^b (2)	0.9754 (4)
s3	0.0000 ^b (1)	0.0002 ^b (3)	0.0000 ^b (2)	0.0007 ^b (4)
s4	0.0000 ^b (1)	0.2452 (4)	0.0005 ^b (2)	0.1714 (3)
s5	0.3185 (4)	0.0034 ^b (2)	0.0000 ^b (1)	0.0218 ^a (3)
s6	0.0036 ^b (2)	0.0000 ^b (1)	0.0946 (4)	0.0223 ^a (3)

^a Significance at the 5% level (paired, two-sided t-test).

^b At the 1% level.

Values in parentheses indicate the ranks of *p*-values from more to less significant.

each subject, for the best combination of number of states and number of mixtures, and for different number of features are reported in figure 6.

Even with only one feature, all subjects performed significantly better than the chance. Some subjects showed a slight decrease in the performance as less features were used, while some showed the opposite tendency. Although the overall performance of all subjects decreased when a limited number of features was used, the overall performance between using only one feature (80.9%) and using all the four features (84.5%) was not significantly different (paired t-test, $p > 0.05$).

4. Discussion

The goal of this study was to investigate the feasibility of detecting motor execution—specifically isometric pinching in index finger/thumb opposition—with a BMI based only on measurements of physiological signals from the ANS. Four physiological features were measured (mean blood pressure, breathing rate, skin conductance response and heart rate) during an isometric pinching task. The acquired physiological signals were used to train

a classifier based on a dual HMM. We hypothesized that activity in cortical areas can be detected by monitoring changes of the ANS, instead of measuring directly at the supraspinal level. We performed an experiment with six healthy subjects the results of which showed that motor execution can be accurately classified based only on peripheral physiological signals.

We hypothesized that such a BMI based on ANS could achieve similar performance in detecting motor execution as a BCI directly based on signals from the central nervous system. This study showed that motor execution can be accurately classified based only on peripheral physiological signals with an accuracy of 84.5%. These results are in line with recent BCI studies that employed EEG to detect movement intention (Boye *et al* 2008) and motor imagery (Tsui *et al* 2009). Few non-invasive BCIs, mainly based on fMRI techniques, have shown higher accuracy levels (Lee *et al* (2010) achieved an accuracy above 90%). However, the infrastructural needs, electromagnetic compatibility limitations and high associated costs make fMRI-based BCIs inappropriate for standard robotic rehabilitation. On the other hand, our results slightly outperformed fNIRS-based BCIs employed to classify mental tasks such as music imagery (Power *et al* 2010) and mental arithmetic (Falk *et al* 2011).

In this study, fNIRS was also employed to simultaneously record brain activity in motor areas (contralateral primary motor cortex and ventral premotor cortex). The brain hemodynamics recorded with fNIRS were employed to train a similar dual HMM classifier (Zimmermann *et al* 2012). The results showed that the classifier based only on the signals measured from the central nervous system with fNIRS achieved an average accuracy of 79.4%, i.e. a slightly lower performance than the classifier based on the ANS response presented here. On the other hand, when the four physiological features described in this paper (mean blood pressure, breathing rate, skin conductance response and heart rate) were added as auxiliary observations into the HMM, the classification accuracy increased significantly to 88.5%. This is in line with recent studies on hybrid BCIs that used brain imaging methods in conjunction with self-paced physiological signals to improve the classification performance (for a review see Pfurtscheller *et al* (2010)). While physiological measures have been successfully employed to improve the accuracy in hybrid BCIs (Falk *et al* 2011, Zimmermann *et al* 2012), the ANS responses have never been employed as the unique information sources to detect motion execution. This paper aims at filling this gap and investigates the feasibility of a BMI to detect motor execution, monitoring only changes in peripheral autonomic signals, reaching similar accuracy levels as hybrid BCIs.

Although recent studies have already introduced the concept of BMIs, where physiological signals can be self-controlled and used to detect functional intent (Blain *et al* 2008), these previous studies are fundamentally based on self-paced physiological signal changes, and thus subjects must be active agents in the changes of their ANS. In our approach, subjects were not requested to change their normal physiological signal responses based on the protocol stimuli. Most of the studies that worked with similar non-self-paced biosignal decoders are found in the field of psychophysiology, where the goal is to estimate subjects' emotions (Kim and Andre 2008), mental workload (Wilson and Russell 2003, Collet *et al* 2009) and activity engagement (Kushki *et al* 2012), instead of function intention.

A recently completed study investigated the use of non-self-paced peripheral autonomic signals to detect music imagery (Falk *et al* 2010). Regardless of the fact that the goal was to decode music imagery instead of motor execution, there are some relevant similarities between these two studies. First, both studies use only physiological autonomic signals (although they used skin temperature, while here mean blood pressure was used). Both studies optimized the number of states and mixtures of a dual HMM classifier and achieved similar accuracy levels (93% in Falk *et al* (2010), and 84.5% here). The smaller accuracy level achieved in our work may be due to the fixed observation window length: they optimized the window lengths per

subject, while we fixed them for all subjects in order to reduce the time required to train the classifier.

HMMs are well known in temporal pattern recognition applications. However, despite their higher ability to classify time-sequential data, compared to discriminative approaches (Sitaram *et al* 2007, Obermaier *et al* 2001), they have barely been used in physiology classification (Kulic and Croft 2007, Falk *et al* 2010). In this paper, we showed that HMMs are a valuable tool to classify motor execution based on time-varying physiological signals. There are, however, some issues with HMMs that must be considered. The per subject optimization of number of states and mixtures requires a large computation time. Here, we studied the effect that fixing the number of states and mixtures for all subjects had on the overall classifier accuracy and found that the optimal fixed model resulted in just a slight reduction of the average classifier accuracy (83.9%). Thus, in order to reduce calibration computation time during normal BCI investigations, it seems reasonable to try to find a reliable fixed HMM structure in future experiments.

Some subjects performed significantly better than others. We also noted that the subset of physiological signals with significant changes was different for each participant. We found a significant correlation between the classifier accuracy and the number of significant features in each subject. We performed feature reduction using statistical tools to test how the reduction of observations affected the classifier performance. We did not find an increase in the accuracy in subjects with a reduced number of significant features. Thus, it was not the inclusion of nonsignificant features what decreased the classifier performance. Interestingly, we did not find a clear accuracy decrease neither when we reduced the number of significant physiological signals used in the classifier. The effect of removing features was dependent on each subject's specific ANS responses.

This finding contradicts recent studies that found a clear monotonic increase in classification performance as more physiological signals were added to the decoder (Falk *et al* 2010, Kushki *et al* 2012). A possible explanation is that different autonomic systems may react in different ways while performing a movement, compared to a more homogenous response to music imagery (Falk *et al* 2010), and activity engagement (Kushki *et al* 2012). Furthermore, the study reported by Kushki *et al* (2012) was performed with individuals with cerebral palsy and muscular dystrophy who presented some physiological differences due to their disabilities (i.e. features related to respiration and the cardiovascular system may have been affected in some subjects). An *a priori* detection of the optimal number of features based on training data, as suggested by Kushki *et al* (2012), could improve the classifier performance for each subject. As an indicative value, selecting the optimum number of features based on all trial data, resulted in an overall performance of 91.6%. However, such an optimization process could also increase the time required to train the classifier.

A major challenge in our research is the comparatively long time periods needed before sufficient information is available to make a decision. As expressed by Blain *et al* (2008), while some EEG-based BCIs have achieved information transfer rates of up to 27.15 decisions min^{-1} , to date BMIs that use only peripheral autonomic signals require at least 30 s to make an accurate detection. In this study, a very conservative observation window length was fixed to 15 s (chosen based on the minimum rest period length). Furthermore, a shift of 5 s was applied to account for physiological signal latencies. This led to a maximum detection delay of 20 s. Although 20 s may be seen as an unreasonable delay for BCI applications, for severely disable individuals who rely on access technologies to move and communicate, speed may not be critical. In a survey of 17 patients in the final stage of ALS who were extensively informed about the possibilities and advantages of an invasive electrode-based BCI, only one agreed to implantation. Patients refused the surgical procedure and preferred the slow non-invasive

system. They argued that time is no issue if one is completely paralyzed (Birbaumer 2006). Priority will be given in further research steps to shorten this relatively long time delay. A possibility could be to select the window lengths optimally for each subject in order to reduce the delay in subjects with faster ANS responses.

The study reported here was conducted in healthy subjects without neurological lesion. We chose to first study healthy subjects in order to evaluate the normative responses of the non-injured ANS during motion execution. Results from this study provide an important starting point and a framework for comparison for future studies with subjects with neurological injury. As presented in this paper, physiological signals vary significantly between subjects. Neurological injuries, such as stroke or spinal cord injury, may affect the autonomic system, which may introduce further variations in the peripheral signals. For example, traumatic brain injury survivors are known to show abnormalities in the autonomic system (hypofunction or hyperfunction) and show an asymmetric sweating with cold hemiplegic limbs that can affect the SCR signal (Korpelainen *et al* 1999, 1993). Patients with complete spinal cord injury showed no changes in electrodermal activity below the level of injury (Cariga *et al* 2002). SCR was shown to be significantly different in patients with multiple sclerosis (Yokota *et al* 1991). On the other hand, some recent studies have shown the feasibility of using some of the physiological signals presented here (i.e. heart rate, SCR and breathing rate) in stroke rehabilitation (Koenig *et al* 2011, Novak *et al* 2011) and with individuals with severe physical disabilities, such as cerebral palsy and muscular dystrophy (Kushki *et al* 2012). Future work with subjects with neurological injuries will focus on determining if the injured ANS can be consistently employed to control a body–computer interface. We speculate that a good classifier accuracy could still be achieved if a physiological signals analysis with patients is performed prior to the training of the classifier. Weak or absent physiological responses can be discarded by means of feature reduction algorithms (Kushki *et al* 2012), similar to the statistical approach used in this paper.

The experiment design also suffers from some limitations. It is well known that attention and mental load significantly affect ANS responses. It is therefore possible that the differences between ‘activity’ and ‘rest’ periods reported here are associated with mental load, instead of motor execution. A well-designed control task is needed (e.g. mental arithmetics, counting backwards) to really conclude that what is classified is in fact motor execution. Moreover, the proposed method was designed to detect motion execution using data from healthy participants who were actively pinching. Motor imagery has been proposed as a strategy to detect motion intention in BCI studies (Falk *et al* 2011, Power *et al* 2010, Tsui *et al* 2009). However, we chose to first study isometric pinching partly for convenience (i.e. it is a simple, well-controlled task that minimizes subject movement), but also because motor imagery does not allow for a systematic assessment of subject’s performance (i.e. motor imagery ability strongly varies among subjects (Sharma *et al* 2006)). It is important to establish the normative mechanisms of the ANS during motion execution, thereby providing a framework for comparison for future studies with motor imagery. Future work will focus on testing with a larger group of subjects to determine if motor imagery yields similar results.

Finally, though physiological signals are easy to measure, they are also affected by different environmental disturbances (e.g. auditory or visual stimuli, external temperature) and by the amount of physical activity. In this study, all these disturbances were minimized by conducting the physiological measurements in a silent, dark room while subjects lay supine. However, such a setup is not realistic in a standard therapeutic environment. Ideally, the use of a wide range of different physiological features could account for these undesirable disturbances. Although physiological signals are prone to habituation, no signal degradation was observed during the experiment described here. A possible explanation is that the random

presentation of sequences and the complex reference force profile constantly engaged subject's active participation. In order to reduce the negative effects of physiological signal habituation, future experimental protocols will be designed to actively engage the subject in an assist-as-needed manner (Zimmerli *et al* 2012). The equipment employed to measure biosignals in this study was selected for convenience (it already existed in our laboratories), but other compact, wireless and easy-to-use solutions exist on the market (e.g. Bluetooth heart rate monitors).

5. Conclusion and outlook

This study showed the feasibility of a BMI to detect motor execution by monitoring only changes of the ANS. Motor execution was accurately classified using a dual HHM classifier based on only peripheral physiological signals with an accuracy level of 84.5%. These results are very encouraging to perform further research on the use of the autonomic system in BMIs for the treatment of severely impaired neurologic patients.

The long term goal of this project is to develop novel human-oriented strategies that enhance the interaction between the robotic system and the user and to incorporate them into robotic systems (e.g. for upper extremity neuro-rehabilitative training). In particular, the robotic system should estimate intention in a continuous manner so that it can optimally assist a human in the anticipated reaching, grasping or manipulation movement. With this approach, participants will control their own movements, while the robotic device will compensate for weakness. The use of physiological signals and binary classifiers may not be enough to achieve the ultimate goal of a continuous decoder. Hence, we plan to use sensor fusion, such that the most likely motor intention can be extracted from a pool of different information sources. These sources include not only physiological recordings, but more sophisticated context analysis (task knowledge and motion history information), gaze and head movement recordings, and recordings of dynamic and kinematic movement components. As an ultimate goal, we plan to incorporate the brain into the loop, integrating measurements of cortical activation acquired through fNIRS.

Acknowledgments

This work was supported by the Swiss National Science Foundation through the National Centre of Competence in Research Robotics and on Neural Plasticity and Repair, and by the CHIRP1 ETH Research grant (CHIRP1-Application CH1-02 09-3, Cortically-Driven Assistance Adaptation during Sensorimotor Training).

References

- Birbaumer N 2006 Brain-computer interface research: coming of age *Clin. Neurophysiol.* **117** 479–83
- Blain S, Chau T and Mihailidis A 2008 Peripheral autonomic signals as access pathways for individuals with severe disabilities: a literature appraisal *Open Rehabil. J.* **1** 27–37
- Boye A T, Kristiansen U Q, Billinger M, do Nascimento O F and Farina D 2008 Identification of movement-related cortical potentials with optimized spatial filtering and principal component analysis *Biomed. Signal Process. Control* **3** 300–4
- Cariga P, Catley M, Mathias C J, Savic G, Frankel H L and Ellaway P H 2002 Organisation of the sympathetic skin response in spinal cord injury *J. Neurol. Neurosurg. Psychiatry* **72** 356–60
- Christov I 2004 Real time electrocardiogram QRS detection using combined adaptive threshold *Biomed. Eng. Online* **3** 28
- Collet C, Averty P and Dittmar A 2009 Autonomic nervous system and subjective ratings of strain in air-traffic control *Appl. Ergon.* **40** 23–32

- Coyle S, Ward T, Markham C and McDarby G 2004 On the suitability of near-infrared (NIR) systems for next-generation brain-computer interfaces *Physiol. Meas.* **25** 815–22
- del R Millan J *et al* 2010 Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges *Front. Neurosci.* **4** 161
- Falk T H, Guirgis M, Power S, Blain S and Chau T 2010 On the use of peripheral autonomic signals for binary control of body-machine interfaces *Physiol. Meas.* **31** 1411
- Falk T, Guirgis M, Power S and Chau T 2011 Taking NIRS-BCIs outside the lab: towards achieving robustness against environment noise *IEEE Trans. Neural Syst. Rehabil. Eng.* **19** 136–46
- Kim J and Andre E 2008 Emotion recognition based on physiological changes in music listening *IEEE Trans. Pattern Anal. Mach. Intell.* **30** 2067–83
- Koenig A, Novak D, Omlin X, Pulfer M, Perreault E, Zimmerli L, Mihelj M and Riener R 2011 Real-time closed-loop control of cognitive load in neurological patients during robot-assisted gait training *IEEE Trans. Neural Syst. Rehabil. Eng.* **19** 453–64
- Korpelainen J, Sotaniemi K and Myllylä V 1999 Autonomic nervous system disorders in stroke *Clin. Auton. Res.* **9** 325–33
- Korpelainen J, Tolonen U, Sotaniemi K and Myllylä V 1993 Suppressed sympathetic skin response in brain infarction *Stroke* **24** 1389–92
- Kulic D and Croft E 2007 Affective state estimation for human-robot interaction *IEEE Trans. Robot.* **23** 991–1000
- Kushki A, Andrews A J, Power S D, King G and Chau T 2012 Classification of activity engagement in individuals with severe physical disabilities using signals of the peripheral nervous system *PLoS One* **7** e30373
- Lee S, Halder S, Kübler A, Birbaumer N and Sitaram R 2010 Effective functional mapping of fMRI data with support-vector machines *Human Brain Mapp.* **31** 1502–11
- Lotze M, Braun C, Birbaumer N, Anders S and Cohen L G 2003 Motor learning elicited by voluntary drive *Brain* **126** 866–72
- Malmivuo J and Plonsey R 1995 *The Electrodermal Response* (New York: Oxford University Press)
- Marchal-Crespo L and Reinkensmeyer D 2009 Review of control strategies for robotic movement training after neurologic injury *J. Neuroeng. Rehabil.* **6** 20
- Muller-Putz G, Scherer R, Brunner C, Leeb R and Pfurtscheller G 2008 Better than random? a closer look on BCI results *Int. J. Bioelectromagn.* **10** 52–5
- Murphy K 1998 Hidden Markov model (HMM) toolbox for Matlab, version 3 (PMTK3) <http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html>
- Novak D, Mihelj M, Zihler J, Olensek A and Munih M 2011 Psychophysiological measurements in a biocooperative feedback loop for upper extremity rehabilitation *IEEE Trans. Neural Syst. Rehabil. Eng.* **19** 400–10
- Obermaier B, Guger C, Neuper C and Pfurtscheller G 2001 Hidden Markov models for online classification of single trial eeg data *Pattern Recognit. Lett. RECPAD2000: Proc. 11th Portuguese Conf. on Pattern Recognition*; **22** 1299–309 (selected paper)
- Perez M A, Lugholt B K, Nyborg K and Nielsen J B 2004 Motor skill training induces changes in the excitability of the leg cortical area in healthy humans *Exp. Brain Res.* **159** 197–205
- Pfurtscheller G, Allison B Z, Bauernfeind G, Brunner C, Escalante T S, Scherer R, Zander T O, Mueller-Putz G, Neuper C and Birbaumer N 2010 The hybrid BCI *Front. Neurosci.* **4** 42
- Power S D, Falk T H and Chau T 2010 Classification of prefrontal activity due to mental arithmetic and music imagery using hidden Markov models and frequency domain near-infrared spectroscopy *J. Neural Eng.* **7** 026002
- Rabiner L 1989 A tutorial on hidden Markov models and selected applications in speech recognition *Proc. IEEE* **77** 257–86
- Rossini P M and Dal Forno G 2004 Integrated technology for evaluation of brain function and neural plasticity *Phys. Med. Rehabil. Clin. North Am.* **15** 263–306
- Scheirer J, Fernandez R, Klein J and P R W 2002 Frustrating the user on purpose: a step toward building an affective computer *Interact. Comput.* **14** 93–118
- Sharma N, Pomeroy V M and Baron J-C 2006 Motor imagery *Stroke* **37** 1941–52
- Sitaram R, Zhang H, Guan C, Thulasidas M, Hoshi Y, Ishikawa A, Shimizu K and Birbaumer N 2007 Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface *Neuroimage* **34** 1416–27
- Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology 1996 Heart rate variability: standards of measurement, physiological interpretation, and clinical use *Circulation* **93** 1043–65
- Tsui C, Gan J and Roberts S 2009 A self-paced brain-computer interface for controlling a robot simulator: an online event labelling paradigm and an extended Kalman filter based algorithm for online training *Med. Biol. Eng. Comput.* **47** 257–65
- van Gerven M *et al* 2009 The brain-computer interface cycle *J. Neural Eng.* **6** 041001

- Wagner J, Kim J and Andre E 2005 From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification *Multimedia and Expo 2005: IEEE Conf. Proc. on (ICME 2005)* pp 940–3
- Wilson G and Russell C 2003 Real-time assessment of mental workload using psychophysiological measures and artificial neural networks *Hum. Factors* **45** 635–43
- Yokota T, Matsunaga T, Okiyama R, Hirose K, Tanabe H, Furukawa T and Tsukagoshi H 1991 Sympathetic skin response in patients with multiple sclerosis compared with patients with spinal cord transection and normal controls *Brain* **114** 1381–94
- Zimmerli L, Krewer C, Gassert R, Muller F, Riener R and Lunenburger L 2012 Validation of a mechanism to balance exercise difficulty in robot-assisted upper-extremity rehabilitation after stroke *J. Neuroeng. Rehabil.* **9** 6
- Zimmermann R, Marchal-Crespo L, Edelmann J, Lamercy O, Fluet M, Riener R, Wolf M and Gassert R 2012 Detection of motor execution using a hybrid fNIRS-biosignal BCI: a feasibility study *J. Neuroeng. Rehabil.* at press
- Zimmermann R, Marchal-Crespo L, Lamercy O, Fluet M, Riener R, Wolf M and Gassert R 2011 Towards a BCI for sensorimotor training: initial results from simultaneous fNIRS and biosignal recordings *EMBC'11: Proc. 33rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* pp 6339–43